Contract N00014-86-K-0173

USING FEASIBILITY CUTS TO ACCELERATE THE CONVERGENCE OF MODES

by

Moshe Eben-Chaime
John J. Jarvis
H. Donald Ratliff

PDRC Report Series 88-02
April 1988

Contract N00014-86-K-0173

USING FEASIBILITY CUTS TO ACCELERATE THE CONVERGENCE OF MODES

by
Moshe Eben-Chaime
John J. Jarvis
H. Donald Ratliff

School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

Contents

1 Introduction **3**

2 Cuts in Networks **4**

3 Bi-Partite Networks and Feasibility Cuts **5**

4 Computational Results **7**
- 4.1 Straightforward Benders' Solution 7
- 4.2 Single-MR Feasibility Cuts 10
- 4.3 Finite Convergence 12

5 Achieving Feasibility **13**
- 5.1 The Best that Can Be Done 13
- 5.2 Achieving Feasibility 15

6 Summary **15**

List of Figures

1 Doing the "Best" 9
2 Single-MR Cuts Added 11
3 Pure Benders' 14

1 Introduction

System for Closure Optimization Planning and Evaluation (SCOPE) is a system for closure planning developed by the Production and Distribution Research Center (PDRC) at Georgia Tech. Two types of decisions are involved in closure planning: 1) allocate the available capacity of different asset types to transportation channel, and 2) load and schedule the various movement requirements to these channels in accordance with the allocated capacities. The problem addressed by SCOPE (and its companion implementation MODES) was formulated (PDRC report 84-09) as a linear program (LP). The LP was decomposed into two sub-problems – channel configuration and movement requirement (MR) assignment. In MODES, channel capabilities are allocated by LIFTCAP, and then serve as (part of the) right-hand side in MRMATE which loads the movement requirements to these channels. The solution method is the row generation procedure of Benders, where the dual variables of MRMATE are used to formulate constraints (cuts, rows) to be added to LIFTCAP.

The specific linear program associated with MRMATE is of the special type known as the minimal cost generalized network flow problem. Its special structure allows fast solution even for very large problems. However, the channel configuration created by LIFTCAP is not guaranteed to allow delivery of all the movement requirements until the final iteration. In many cases finding a feasible solution, i.e. channel configuration which allows 100% delivery, when one exists, is as hard as finding the optimal solution. The network of MRMATE is examined and special classes of cut-sets (partitions of the node set of the network) is identified in this network which are themselves associated with necessary conditions for the network to admit feasible flow. Some of these conditions are then added to LIFTCAP as additional constraints in order to make the channel configuration produced by this sub)problem closer to feasibility. These cuts are thus termed "feasibility" cuts. (KR)

Computational tests were performed to test the efficiency of this approach, and determine the most effective cuts. The main result is that one of the test problems, 123DF01, supplied by USTRANSCOM can be shown to be infeasible in a very few number of iterations compared to previous implementations of Benders procedure. Twenty cuts were added to LIFTCAP, less than 10% of the original 275 rows LIFTCAP consists of in this

instance.

The two-phase simplex method of linear programming is used to solve LIFTCAP. The information of the final phase 1 tableau, when infeasibility is concluded, is used to gain additional insights into ways to modify the problem to achieve feasibility. First, the information is used to determine good ways to relax those feasibility cuts that could not be satisfied. Second, it is used to suggest what can be done in order to make the problem feasible, and allow delivery of all the movement requirements within the specified time horizon.

2 Cuts in Networks

A network is defined by a set of nodes, N, and a set of directed arcs, A. An arc is defined by an ordered pair (i, j), where the tail of the arc is attached to the node i, and its head points to the node j. Of course, i, and j are both members of the node set N. A weight is associated with each arc in A. This weight may be the travel time from i to j, or the cost of shipping single unit of good from i to j. In both cases, this weight may take different values in the opposite direction, that is from j to i. The problem discussed here is the minimal cost network flow problem in which the weights are unit shipping costs, and thus denoted by c_{ij}. Also, in the minimal cost network flow problems, there is a number b_i associated with each node i in the node set N. The quantity b_i indicates the demand or, if negative, the available supply at node i in number of units. The problem is then to find the set of flows x_{ij} that is feasible, i.e. non-negative, and satisfies the conservation of flow equations:

$$\sum_{(i,j)\epsilon A} x_{ij} - \sum_{(j,k)\epsilon A} x_{jk} = b_j \; for \; all \; j\epsilon N, \tag{1}$$

and minimize the total cost:

$$\sum_{(i,j)\epsilon A} c_{ij}x_{ij}. \tag{2}$$

A cutset (cut) in a network is defined when the node set N is partitioned into two disjoint sets, say X and $\bar{X}$. Correspondingly, the arc set A is partitioned into four disjoint sets: $\{(X, X)\}$, $\{(\bar{X}, \bar{X}\}$, $\{(\bar{X}, X)\}$, and

$\{(X,\bar{X})\}$, where $\{(U,V)\}$ denote the set of all arcs such that $i\epsilon U$, and $j\epsilon V$. The last set $\{(X,\bar{X})\}$, is usually called the cutset.

Cuts are very useful in many network problems, and will be used here to accelerate the convergence of MODES.

3 Bi-Partite Networks and Feasibility Cuts

In many cases the network defined by the particular arc set A, over the node set N, has special structure. One such special structure is obtained when the arcs incident to each node are either all directed into it, or all emanating from it. In this case, the node set N may be partitioned into two disjoint sets, say S and D, such that if an arc (i,j) belongs to A, then $i\epsilon S$, and $j\epsilon D$. Assuming each node in N has at least one arc incident to it, the union of S and D is N itself. The network is then said to be bi-partite.

The conservation of flow equations for the bi-partite network may be written separately for the destination set D, and for the set of resources S. Further, the equations for the resource set S can be stated as less then or equal inequalities. That is (1) is replaced by (3), and (4):

$$\sum_{j:(i,j)\epsilon A} x_{ij} \leq s_i \ for\ all\ i\epsilon S \tag{3}$$

$$\sum_{i:(i,j)\epsilon A} x_{ij} = d_j \ for\ all\ j\epsilon D. \tag{4}$$

This, plus possibly bounds on the flows $-x_{ij}$, is exactly the formulation of the pure network model of MRMATE (see PDRC 84-09, and PDRC 85-03).

A special class of cuts in bi-partite networks is defined when in the partition $(X,\bar{X})$, $\bar{X}$ is restricted to be a subset of D (including D itself). One such cut, Y, a subset of X may be defined to be the set of all nodes i for which there exist an arc (i,j) in the cutset for some j:

$$Y = \{i:\ (i,j)\epsilon\{(X,\bar{X})\},\ for\ at\ least\ one\ j\}. \tag{5}$$

An immediate observation is that any demand in the set $\bar{X}$ can be satisfied only by the resources in the set Y. This implies the necessity

of the following condition for any bi-partite network to admit a feasible solution:

$$\sum_{i \epsilon Y} s_i \geq \sum_{j \epsilon \bar{X}} d_j \ for\ any\ (X, \bar{X})\ s.t.\ \bar{X}\ is\ a\ subset\ of\ D. \tag{6}$$

This condition simply says that since the resources in Y are the only available to satisfy demand in $\bar{X}$, the total supply available from the set Y, should be no smaller than the total demand created by $\bar{X}$. Condition 6 holds for any subset $\bar{X}$ of D, including D itself. Unfortunately there are too many cuts in this class, yet, if the network is highly structured only few of them are really needed to ensure feasibility.

In MODES, the resources are the channels, and hence the single index i used before is replaced by combination of a, i, j, t which indicate the asset, POE, POD, and time period associated with the channel. The set D comprises the movement requirements indexed by r. The specific formulation of (3) and (4) for the pure model is thus given by (7), and (8):

$$\sum_r x_{raijt} \leq N_{aijt} c_{aijp}\ for\ all\ a, i, j,\ and\ t \epsilon T_p \tag{7}$$

$$\sum_{aijt} x_{raijt} = M_r\ for\ all\ r. \tag{8}$$

The sets $\bar{X}$ are formed by taking subsets of the movement requirements, and the associated sets Y are all the MRMATE time expended channels which can deliver the subset of MR's being considered. To be more precise:

$$\sum_{aijt:\ (r,aijt) \epsilon A\ for\ some\ r \epsilon \bar{R}} N_{aijt} c_{aijp} \geq \sum_{\bar{R}} M_r, \tag{9}$$

which implies that, at least, a potential exists to deliver the movement requirements. The only variables in constraint (9) are the channel capacities c_{aijp}, which are determined by LIFTCAP. These remain fixed when MRMATE is being solved. Since MRMATE will admit a feasible solution only if (9) is satisfied for all subsets $\bar{R}$, (9) must also be satisfied by any feasible solution of the overall problem. Hence, equations of this form can be added as constraints to LIFTCAP. These constraints will usually restrict the feasible region of LIFTCAP "cutting" out those solutions that are infeasible in MRMATE, and are therefore termed feasibility cuts.

In most cases the generalized network model is used to model MRMATE. In this model the constraints (7) are replaced by (10):

$$\sum_{r} V_{raijt} x_{raijt} \leq N_{aijt} c_{aijp} \; for \; all \; a, i, j, \; and \; t \epsilon T_p. \tag{10}$$

In general, the constraints (9) can no longer be formed unless each set $\bar{R}$ contains a single movement requirement, in which case the resulting constraints are not so effective. In many cases however even though MRMATE might be originally formulated as a generalized network problem, it can very easily be converted into a pure network model. Further, even in those cases that are indeed generalized networks, more effective cuts can still be found. As an example, cuts can be established for those MRs connected only to sea channels or only to air channels. In this case the multiplier V_{raijt} is actually V_r for any a, i, j, and t.

The constraints (9) are generated only once before the solution procedure starts, and are added to any LIFTCAP problem being solved at every iteration. The computer can easily and automatically generate the single-MR cuts as part of the initialization step. Even for more complicated cuts only a 'onetime' effort is needed. This approach was implemented and tested on the IBM 4341 at Georgia Tech, and the results are reported in the next section.

4 Computational Results

The data set used for testing the feasibility cuts approach was 123DF01, or RID1. This data set does not converge with straightforward application of Benders' procedure. The results of this straightforward application are reported here to form abase-line for comparison. Feasibility cuts are then added, first in the single-MR form, and then more complicated and tight cuts were formulated.

4.1 Straightforward Benders' Solution

Benders' method starts with a feasible solution of LIFTCAP, for example, the maximal throughput solution described in PDRC 85-03. With this start-up solution, 18 iterations of the algcrithm were performed on RID1

before storage limitations were exceeded. The best MRMATE objective value, which serves as an upper bound on the overall objective, is $195x10^6$, and is achieved at the same iteration in which the highest percentage of the total MR, 61%, is delivered. The best LIFTCAP objective value, a lower bound on the overall objective, is $-69x10^6$. The negativity of the lower bound as well as the low delivery percentage indicates that demonstration of convergence is not close. Solution progress is shown by the convergence chart of Figure 1. The upper bound, UB, line (solid) is the best MRMATE solution found thus far, while the dashed line shows how MRMATEs objective value fluctuates across iterations. The same holds for the solid and dashed %Delivered lines. The LB line is generated by objective values of LIFTCAP, and is the lower bound of the system optimum.

Two notes on this figure:

1. MRMATE objective values and the delivery percentage are highly correlated, and in Figure 1 the %Delivered dashed line is almost a mirror image of the UB dashed line. This is a result of the artificial arcs with very large (artificial) costs being used to actually solve MRMATE.

2. The behavior of these two dashed lines is highly unstable as indicated by the big jumps of the lines.

Another comparison can made to the solution achieved by the "advanced start" procedure. This procedure opens channels and allocates capacities in order to deliver as much as possible of the movement requirements while maintaining LIFTCAP feasibility. The channel configuration generated by this procedure is thus feasible in LIFTCAP, yet it is not necessarily a basic solution in the LP sense. Consequently it cannot be directly used as an advanced start for Benders. It is a single-pass heuristic solution for the overall problem. When applied to RID1, a channel configuration was produced that is able to deliver 80.5% of the MRs at total cost of $110x10^6$. This is an upper bound on the overall objective, and is much smaller than that created by Benders. Also, this "advanced start" solution delivers about 20% more of the total MRs.

Both solutions are not satisfactory since they do not allow any strong conclusions to be drawn about the optimal solution. Nothing can be said as to what the true objective value is, and how far from it is the $110x10^6$

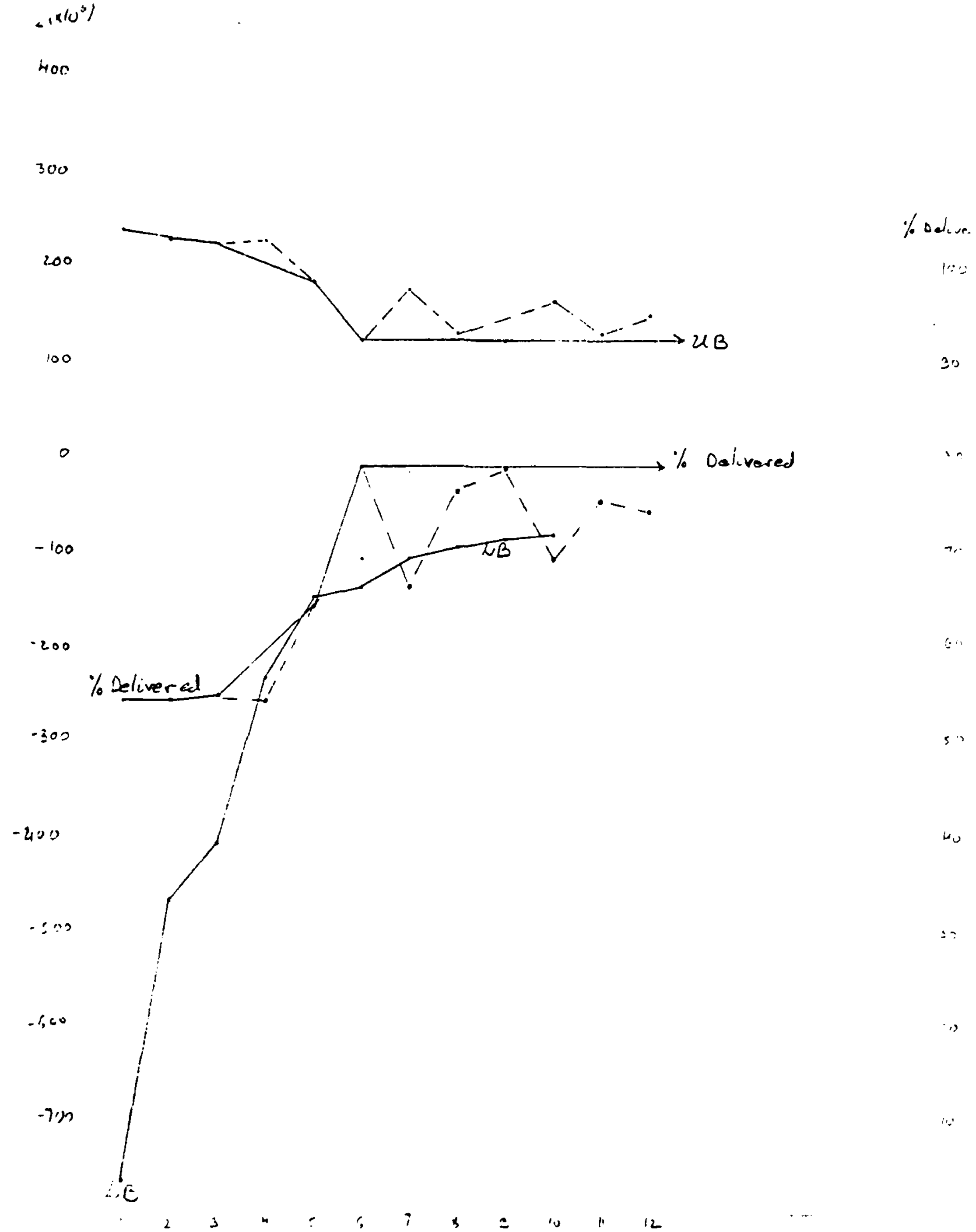

Figure 1: Doing the "Best"

upper bound. Furthermore, it is not even known whether this problem admits a feasible solution, i.e. whether all the MRs be delivered within the specified time horizon.

4.2 Single-MR Feasibility Cuts

To get a rough estimate of the effect of the feasibility cuts, the single-MR cuts were added to LIFTCAP. In this form, each of the sets $\bar{R}$ in (9) contains a single movement requirement. There is one such set, namely one cut, for each of the movement requirements. This results in an addition of 113 constraints to LIFTCAP – approximately 40% of the original 275 rows of this problem (233 network constraints and 42 side constraints). Despite this significant change in problem size, the solution time was in some cases shorter, and never significantly longer.

The computational results demonstrate the benefits of the approach. The first solution of the straight-forward application delivers only 0.19% of the MRs while the corresponding solution with the single-MR cuts begins with 54% delivered. The first upper bound generated by this application is $233x10^6$, and the lower bound is $-765x10^6$. Both of these values are about 1/2 of the numbers generated by the straight-forward application. Comparison of the convergence chart of the single-MR cuts application shown in Figure 2, with that of Figure 1, reveals higher stability in system behavior. This results in a more uniform progress toward convergence.

This application terminated after 12 iterations, delivering 78.2% of the movement requirements at a total cost (upper bound) of $120x10^6$. These results are much better than those of the straight-forward application and are very close to those of the "advanced start" heuristic.

Recall that the addition of the single-MR cuts enlarged LIFTCAP size about 40%. Yet, these cuts are highly redundant. If, for example, two MRs are always connected to the same channels, it suffices to satisfy constraint (9) associated with the largest MR. The constraint associated with the second MR is automatically satisfied by this solution. It then follows that at least some of the single-MR cuts need not be explicitly stated.

This observation, accompanied with the promising results of the single-MR application, motivate careful study of the network of this instance with the aim of generating more effective cuts.

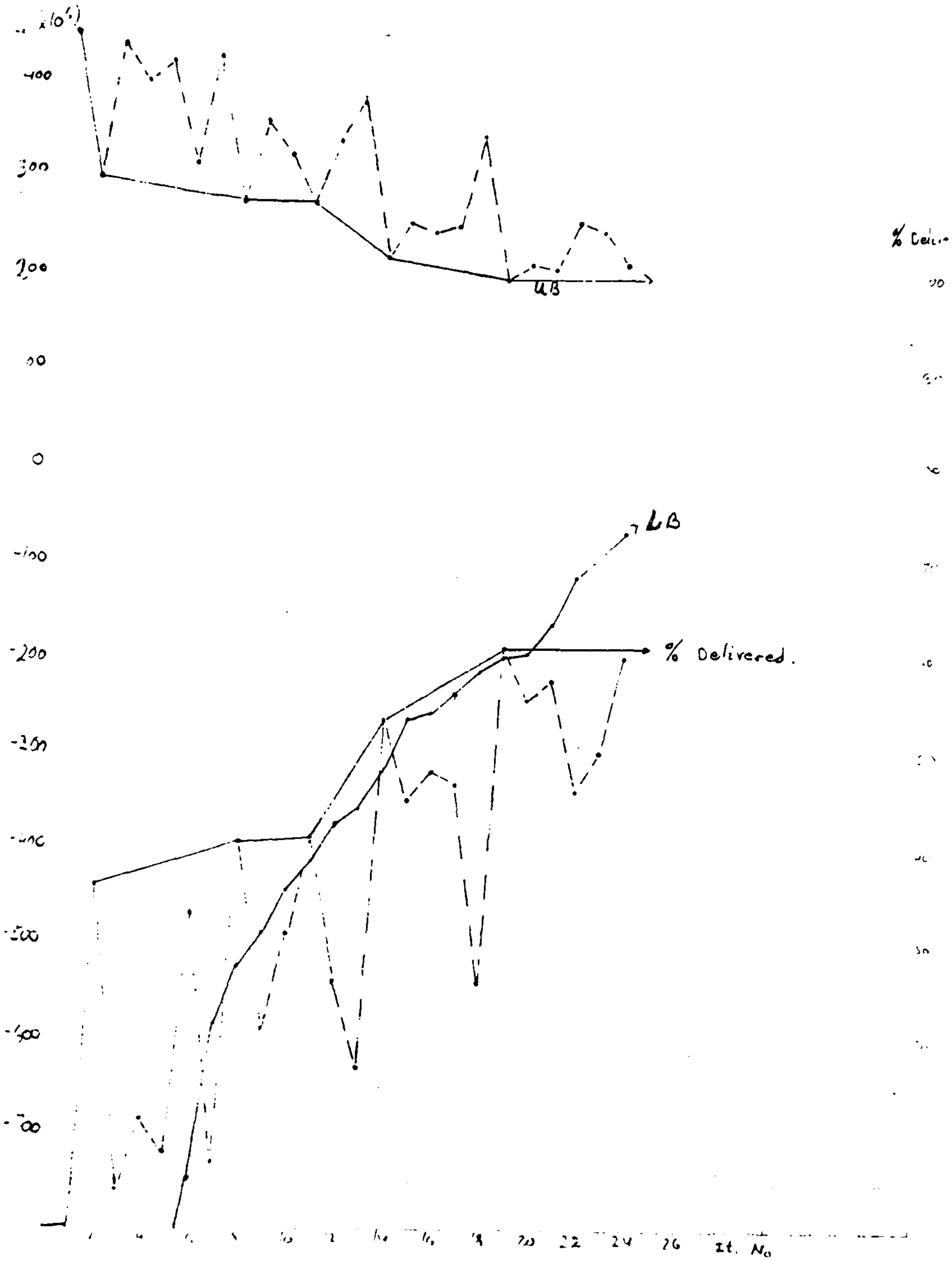

Figure 2: Single-MR Cuts Added

4.3 Finite Convergence

Examination of the network of RID1 indicated the following two properties:

1. The problem can be converted to a pure network. The mode decision for each MR had already been made and therefore each MR is either a sea MR or an air MR. Consequently, the indices a, i, j, and t can be omitted from the multipliers V_{raijt} giving V_r. Transforming the variables of MRMATE by

$$Y_{raijt} = V_r x_{raijt} \textit{ for all } (r, aijt) \textit{ in } A, \tag{11}$$

   and multiplying each right-hand side of (8) by the corresponding V_r, results in an equivalent problem which is in pure network form.

2. The MRMATE problem is separable. First, by air and sea. Second, even in the separated sea (air) networks, groups of MRs-channels that form separated sub-networks of their own could have been identified.

The first property allows the generation of multi-MR cuts. The second suggests how to formulate them – a cut for each sub-network – forcing the delivery potential of the sub-network to be no smaller than the total requirement of this network. As a result, 12 sub-network cuts were generated.

Further investigation of the network revealed that 86 out of the 113 movement requirements are all associated with one of the sea sub-networks; and among those many are grouped together and always connected to the same, usually disjoint, subsets of channels. Eight additional cuts were created for these groups. These 20 cuts seem to dominate most of the single-MR cuts. Hence, the 113 single-MR cuts were all replaced by the newly created 20 cuts. LIFTCAP then had 295 rows, about the same as its original size. Moreover, analysis of the new 20 cuts indicates that a feasible solution, if one exists, will enable delivery of at least 90% of the total MRs!

In the next section, it will be shown how the cuts used to prove infeasibility can be use to: 1) help in finding the best that can be done, and 2) indicate what should be done to make the problem feasible.

5 Achieving Feasibility

LIFTCAP is formulated as a linear program (LP), and it is solved using the two-phase simplex method. This method first attempts to find a feasible solution to the problem during phase 1. Upon success, phase 2 finds the solution which is not only feasible but also optimal. In phase 1, even when infeasibility occurs, the information provided by the final tableau may still be useful. One way to use it is by relaxing those constraints that can not be satisfied. Another way is to modify the problem so as to achieve feasibility. The results of applying these two approaches to RID1 are reported in this section.

5.1 The Best that Can Be Done

Only 3 out of the 20 feasibility cuts added to LIFTCAP could not be satisfied. These were the twelfth sub-network cut, the one associated with 86 of the MRs, and two of the cuts associated with subsets of these 86 MRs. In this case, it is enough to relax the sub-network cut, and only one of the other two to make the problem feasible. The constraints were relaxed by decreasing their right-hand side, since they are all 'greater than or equal' constraints. The amount by which the right-hand side is decreased is the minimum required, i.e. the value of the artificial variable associated with the constraint in the final phase 1 tableau.

The values of the two artificial variables under consideration, indicate that the best that can expected is delivery of somewhere between 80% (the 'advanced start' solution) and 85% of the MRs. The exact value is obtained only upon finite convergence of MODES. Yet the results obtained when the above relaxation was implemented are close enough to conclude that the true value is about 84-85%.

This application terminated after 13 iteration. However, the best delivery percentage, 84.4%, as well as the best MRMATE objective value, $91.7x10^6$, were obtained after 3 iterations, and did not essentially change thereafter (see Figure 3). Most of the time was then used to bring the lower bound closer to the upper bound. For the first time there is a positive lower bound for this dataset. Table 1 compares the various solutions reported herein.

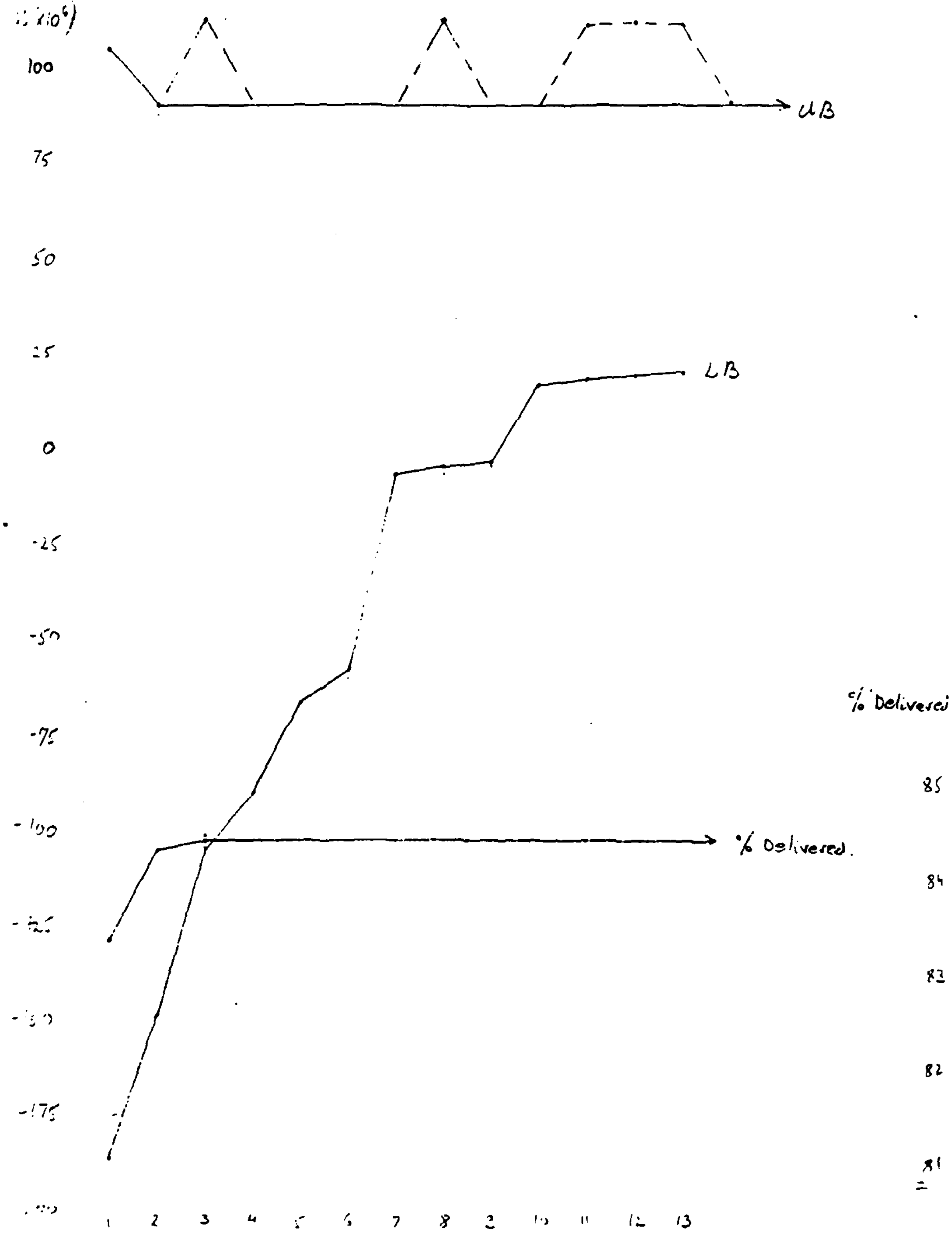

Figure 3: Pure Benders'

Table 1: Solution Comparisons

Application	upper bnd ($x10^6$)	lower bnd ($x10^6$)	% del	# iter
straight fwrd.	195	-69	61	18
Single-MR cuts	120	-84	78.2	12
'Adv start'	110	–	80.5	–
20 cuts relaxed	Infeasible			–
20 cuts	91.7	20	84.4	13

5.2 Achieving Feasibility

Relaxing the cuts, as in the previous section, makes the problem feasible only in the mathematical sense. In order to make it practically feasible, the movement requirements and time expanded channels (MRMATEs nodes) associated with the violated cuts should be identified. Then, a decision should be made as to whether the size of the corresponding MRs should be reduced – extending the time horizon. Or a way might be found to increase channel capabilities, i.e. increase the right-hand side of those original LIFTCAP constraints that 'cause' the infeasibility. In RID1, by increasing the total volume of sea assets available during LIFTCAP period 2, and 3 by approximately 25%, delivery of at least 99.3 percent is achieved (by the 'advanced start'). Benders' method under these conditions delivers 95% right as its starting solution. Alternatively, two movement requirements with volume of about 42,000 tons (19% of the total volume) could have been removed from the problem (delivered later in time). Benders again begins by delivering 95% of the remaining MRs.

6 Summary

A methodology has been proposed to accelerate the convergence of MODES, by speeding up the process of finding a feasible solution or determining infeasibility. In the latter case the final phase 1 tableau can be used to identify the best that can be done, and/or what should be done in order to make the problem feasible. The computational experience has indicated infeasibility in less than a full iteration of the solution procedure.

Finally, it should be noted that the cuts developed in this report can be shown to be valid cuts within the framework of Benders' method used in MODES. Hence the proposed acceleration technique is consistent with this system.